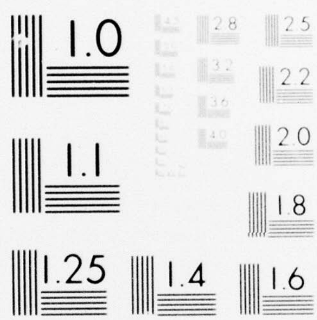
AD-A046 603 MARYLAND UNIV COLLEGE PARK DEPT OF MATHEMATICS F/G 12/1
SUFFICIENCY AND THE NUMBER OF LEVEL CROSSINGS BY A STATIONARY P--ETC(U)
AUG 77 B KEDEM AFOSR-77-3332
UNCLASSIFIED TR77-41 AFOSR-TR-77-1259 NL

| OF |
AD
A046603

END
DATE
FILMED
12-77
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AFOSR-TR- 77-1259

(5)

AD A046603

SUFFICIENCY AND THE NUMBER OF LEVEL CROSSINGS BY A STATIONARY PROCESS

by

Benjamin Kedem

DEPARTMENT OF MATHEMATICS
University of Maryland
College Park, Maryland
20742

MD77-52-BK
TR77-41
August 1977

SUFFICIENCY AND THE NUMBER OF LEVEL CROSSINGS BY A STATIONARY PROCESS

Benjamin Kedem
University of Maryland, College Park

Summary. It is shown how to derive the exact distribution of the number of axis crossings by a stationary process when the binary process obtained by clipping the original process is a pth-order Markov chain. The same method is used in deriving the asymptotic distribution of the number of upcrossings of a high level by a stationary process.

Key words and phrases: binary, Markov chain, level crossings, symbol changes, upcrossings, high level

1. Introduction. Let Z_t, $t=1,\ldots,n$, be a strictly stationary time series, and let X_t, $t=1,\ldots,n$, be the binary time series which takes the values 1 whenever $Z_t \geq a$ and 0 otherwise. X_t as well as quantities defined by it should be indexed by the level a, but except in one case we shall avoid this indexing for the sake of simplified notation. Associated with X_t are the statistics

$$D(n) = 2\sum_{t=1}^{n} X_t - 2\sum_{t=2}^{n} X_t X_{t-1} - (X_1 + X_n) \quad \text{and} \quad D_a(n) = \sum_{t=1}^{n} X_t - \sum_{t=2}^{n} X_t X_{t-1}.$$

$D(n)$ counts the number of symbol changes in the binary series and hence it counts the number of crossings of level a by Z_t. When $X_1 + X_n = 0$, $D_a(n)$ counts the number of upcrossings of level a by Z_t. We shall find the distribution of $D(n)$, n fixed, for level $a = 0$ and the asymptotic distribution of $D_a(n)$, as $a, n \to \infty$ in a suitable manner, when X_t is either a first or second-order Markov chain. The same technique applies to higher order chains.

We shall make use of the results in Kedem (1976,a). Consequently we define

$$p = P_r(Z_t \geq a), \quad \lambda_k = P_r(Z_t \geq a | Z_{t-k} \geq a), \qquad k=1,2$$

$$\mu = P_r(Z_t \geq a | Z_{t-1} \geq a, \ Z_{t-2} \geq a),$$

$S = \Sigma X_t$, $R_1 = \Sigma X_t X_{t-1}$, $R_2 = \Sigma X_t X_{t-2}$, $C = \Sigma X_t X_{t-1} X_{t-2}$, $H = X_1 + X_n$, $U = X_2 + X_{n-1}$, $V = X_1 X_2 + X_{n-1} X_n$.

For a review of level crossings problems and an extensive bibliography see Leadbetter (1972).

2. The number of axis crossings. In this section $a = 0$ and $p = 1/2$. That is $P_r(Z_t \geq 0) = 1/2$.

Theorem 1. If X_t is a first-order Markov chain, then the number of axis crossings by Z_t, $t=1,\ldots,n$, has a binomial distribution $b(n-1, 1-\lambda_1)$.

Proof. The probability of a 0-1 series for which $D(n) = d$ is given by

$$P_r(X_1 = x_1, \ldots, X_n = x_n) = \frac{1}{2} (1-\lambda_1)^d \lambda_1^{(n-1)-d} \tag{1}$$

and there are $2\binom{n-1}{d}$ such sequences. Multiply this number by (1) to obtain the desired binomial distribution.

Observe that under the conditions of the theorem $D(n)$ is minimal sufficient for λ_1 and the maximum likelihood estimate of λ_1 is $\hat{\lambda}_1 = \{(n-1)-D(n)\}/(n-1)$ while $\sqrt{n}(\hat{\lambda}_1 - \lambda)$ is asymptotically $N(0, \lambda_1(1-\lambda_1))$.

Just when may we expect the above binomial distribution to be a reasonable approximation to the actual distribution of the number of axis crossings? So, consider a stationary AR(1) process $Z_t = \phi Z_{t-1} + u_t$, $|\phi| < 1$, u_t are independent $N(0,1)$ variates. For each of 19 values of ϕ 1000 time series of size 5 were generated. The size was fixed at 5 to allow the expected number of successes in each cell in a multinomial experiment to exceed 1 in 1000 repetitions. We wish to test $H_0: D(n) \sim b(n-1, 1-\lambda_1)$ where now $n = 5$ and $\lambda_1 = \frac{1}{2} + \frac{1}{\pi} \sin^{-1}(\phi)$. The results of the 19 chi-square goodness of fit tests are summarized in Table 1. It is seen that the results are very satisfactory for $-0.588 < \phi \leq 0.600$ where H_0 is accepted at level of significance 0.01. This example indicates that the above binomial distribution is reasonable when neighboring observations in the Z_t series are at most moderately correlated.

Table 1: Observed (expected) frequencies of the number of axis crossings by $Z_t = \phi Z_{t-1} + u_t$, $t = 1,\ldots,5$, u_t N(0,1), in 1000 independent realizations.

ϕ	λ	0	1	2	3	4	$\chi^2_{(4)}$
.809	.800	457	322	171	43	7	56.235*
		(409.6)	(409.6)	(153.6)	(25.6)	(1.6)	
.707	.750	358	366	190	73	13	50.716*
		(316.4)	(421.87)	(210.94)	(46.87)	(3.91)	
.600	.705	259	405	243	80	13	6.505
		(247)	(413.4)	(260)	(72.3)	(7.6)	
.588	.700	253	400	252	81	14	6.303
		(240.1)	(411.6)	(264.6)	(75.6)	(8.1)	
.500	.666	209	391	290	87	23	11.549
		(197.4)	(394.9)	(296.2)	(98.72)	(12.3)	
.400	.631	169	365	318	119	29	7.387
		(158.5)	(370.8)	(325.2)	(126.8)	(18.5)	
.309	.600	143	331	336	157	33	4.483
		(129.6)	(345.6)	(345.6)	(153.6)	(25.6)	
.250	.580	127	315	344	173	41	5.783
		(113.1)	(327.8)	(356)	(171.2)	(31.1)	
.100	.532	85	281	363	215	56	1.908
		(80)	(282)	(372)	(218)	(48)	
.000	.500	64	254	378	233	71	2.436
		(62.5)	(250)	(375)	(250)	(62.5)	
-.100	.468	53	217	366	374	90	2.099
		(48)	(218)	(372)	(282)	(80)	
-.250	.420	37	175	345	315	128	4.007
		(31.1)	(171.2)	(356)	(327.8)	(113.1)	
-.309	.400	30	160	341	326	143	3.580
		(25.6)	(153.6)	(345.6)	(345.6)	(129.6)	
-.400	.369	22	133	335	335	175	6.433
		(18.5)	(126.8)	(325.2)	(370.8)	(158.5)	
-.500	.333	18	113	294	365	210	7.791
		(12.3)	(98.72)	(296.2)	(394.9)	(197.4)	
-.588	.300	18	97	258	358	269	28.779*
		(8.1)	(75.6)	(264.6)	(411.6)	(240.1)	
-.600	.295	18	93	259	358	272	30.116*
		(7.6)	(72.3)	(260)	(413.4)	(247)	
-.707	.250	12	71	213	343	361	50.268*
		(3.91)	(46.87)	(210.94)	(421.87)	(316.4)	
-.809	.200	10	37	157	305	491	92.191*
		(1.6)	(25.6)	(153.6)	(409.6)	(409.6)	

*Indicates that the hypothesis H_0 is rejected at level of significance 0.01.

A more realistic assumption is that X_t displays a higher order dependence. The extension of Theorem 1 to the case when X_t is a k th-order Markov chain is somewhat more involved but straightforward. For this purpose let us consider the second-order case in detail; the k th-order case follows an identical argument.

When X_t is a second-order chain it was shown in Kedem (1976,a) that $\{S,R_1,R_2,C,H,U,V\}$ is a set of sufficient statistics for λ_1,λ_2,μ, and their joint distribution is given there. An equivalent but a more convenient set of sufficient statistics is $\{S,D(n),F,Z',H,U,V\}$ where $D(n) = 2S-2R_1-H$, $F = R_1-C$ is the number of 1-runs in the X_t series with two or more 1's and $Z' = R_2-C$ is the number of 0-runs between the first and last 1 with exactly one 0. It follows that the joint distribution of the last set can be obtained from that of the first one. We have

$$g(s,d,f,z',h,u,v) = P_r(S=s,\ D(n)=d,\ F=f, Z'=z',\ H=h,\ U=u,\ V=v)$$

$$= N(s,d,f,z',h,u,v)K_n(\xi_1\xi_2\xi_3\xi_4)^s(\xi_2\xi_3\xi_4)^{-\frac{1}{2}d}\ (\xi_3\xi_4)^{-f}\xi_3^{z'} \tag{2}$$

$$\cdot\ [(\xi_2\xi_3\xi_4)^{-\frac{1}{2}}\xi_5]^h\xi_6^u\xi_7^v,$$

where $K_n,\xi_1,\xi_2,\ldots,\xi_7$ are functions of $p = 1/2,\lambda_1,\lambda_2,\mu$ and are given in Kedem (1976) and

$$N(s,d,f,z',h,u,v)$$

$$= \binom{2}{\max(h,u)}\binom{\max(h,u)}{v}\binom{\frac{1}{2}(d+h)-1}{z'}\binom{n-s-\frac{1}{2}(d-h)-2}{\frac{1}{2}(d-h)-z'-u+v}\binom{\frac{1}{2}(d-h)}{f-v}\binom{s-\frac{1}{2}(d+h)-1}{f-1}$$

<u>**Theorem 2**</u>. If X_t is a second order Markov chain then the distribution of the number of crossings by Z_t, $t=1,\ldots,n$ is given by

$$P_r(D(n)-d) = \sum_{(h,u,v)}\ \sum_{s=h+u}^{h+u+n-4}\ \sum_{f=v}^{v+\frac{1}{2}(d-h)}\ \sum_{z'=0}^{\frac{1}{2}(d+h)-1} g(s,d,f,z',h,u,v), \tag{3}$$

where (h,u,v) takes values in $(1,2,1),(1,1,1),(1,1,0),(1,0,0)$ when d is odd and in $(2,2,2),(2,1,1),(0,2,0),(2,0,0),(0,1,0),(0,0,0)$ when d is even.

In principle it is possible to extend our method to obtain the distribution of $D(n)$ when the 0-1 series displays a higher order dependence but the joint distribution of the sufficient statistics bebomes messier.

3. Upcrossings of a high level. In this section we shall elicit the Poisson nature of the upcrossings of a high level a by Z_t, by using the above method of examining the joint distribution of several sufficient statistics. The Poisson nature of these upcrossings [3] has been known for nearly twenty years for continuous parameter Gaussian processes under various moment conditions. Z_t, however, is not necessarily Gaussian.

Theorem 3. Assume X_t is a first order Markov chain. If $a,n \to \infty$ such that

(i) $nP_r(Z_t \geq a) = \alpha$, α remains constant,

(ii) $P_r(Z_t \geq a | Z_{t-1} \geq a) = \lambda_1(a) \to \lambda_1$,

then

$$\lim_{a\to\infty} P_r(D_a(n)=k) = \frac{e^{-\alpha(1-\lambda_1)}[\alpha(1-\lambda_1)]^k}{k!}, \qquad k=0,1,\ldots . \tag{4}$$

Proof. A simple combinatorial argument shows that

$$P_r(S=s, D_a(n)=k, H=0) = \binom{s-1}{s-k}\binom{n-s-1}{k} p^k q^{s-n+2} \lambda_1^{s-k}(1-\lambda_1)^{2k} \cdot (1-2p+\lambda_1 p)^{n-1-s-k}.$$

Replace p by α/n and q by $1-\alpha/n$ and note that $\{H=0\}$ becomes a sure event as $a \to \infty$. Then

$$\lim_{a\to\infty} P_r(S=s, D_a(n)=k) = \frac{[\alpha(1-\lambda_1)]^k e^{-\alpha(1-\lambda_1)}}{k!}\binom{s-1}{k-1}(1-\lambda_1)^k\lambda_1^{s-k}, \qquad (5)$$

and sum over s.

As consequences we have firstly

$$\lim_{a\to\infty} P_r(\max_{t=1,\ldots,n} Z_t \le a) = e^{-\alpha(1-\lambda_1)}, \qquad (6)$$

and secondly, the asymptotic distribution of S, the total time spent above a high level a, is the Polya-Aeppli distribution obtained by summing (5) over k, with mean α and variance $\alpha(1+\lambda_1)/(1-\lambda_1)$.

Similar results can be obtained for the second-order case. To simplify matters assume $Z' \to 0$ as $a \to \infty$ with probability one, which happens if and only if $\lambda_2-\lambda_1\mu \to 0$, $a \to \infty$.

Theorem 4. If X_t is a second-order Markov chain such that (i) and (ii) above hold and

(iii) $P_r(Z_t \ge a \mid Z_{t-1}\ge a, Z_{t-2}\ge a) = \mu(a) \to \mu$,

(iv) $(\lambda_2-\lambda_1\mu)^{Z'} \to 1$ with probability one, as $a \to \infty$,

then $D_a(n)$ has an asymptotic Poisson distribution with parameter $\alpha(1-\lambda_1)$.

Proof. From (2) with $p = \alpha/n$ and the fact that $\{H=0, U=0, V=0\}$ becomes a sure event, it follows that

$$\lim_{a\to\infty} P_r(S=s, D_a(n)=k, F=f) = \frac{\alpha^k}{k!} e^{-\alpha(1-\lambda_1)}\binom{k}{f}\binom{s-k-1}{f-1}$$

$$\lambda_1^f \mu^{s-k-f}(1-\mu)^{2f}[1-\lambda_1(2-\mu)]^{k-f}.$$

But

$$\sum_{s=k+f}^{\infty}\binom{s-k-1}{f-1}\mu^{s-k-f} = (1-\mu)^{-f}$$

and

$$\sum_{f=0}^{k} \binom{k}{f} [\lambda_1(1-\mu)]^f [1-\lambda_1(2-\mu)]^{k-f} = (1-\lambda_1)^k ,$$

so that $P_r(D_a(n)=k) \to e^{-\alpha(1-\lambda_1)} [\alpha(1-\lambda_1)]^k/k!$.

4. Some applications.

When parameters of interest are related in some fashion to the number of axis crossings, Theorem 1 can be used in deriving appropriate estimators and their approximate distributions. We bring two such cases.

Estimation in AR(1). Suppose $Z_t = \phi Z_{t-1} + u_t$ is a stationary AR(1) process as above, and suppose it is clipped at level D. If the clipped process X_t approximates a first order Markov chain, then the maximum likelihood estimate of ϕ based on the clipped data is

$$\hat{\phi} = \phi(\hat{\lambda}_1) = \sin \pi \left\{ \frac{(n-1) - (\#\text{ of axis crossings})}{n-1} - \frac{1}{2} \right\}$$

Experience shows [2] that this estimator behaves remarkably well even when $|\phi|$ is close to 1. When $|\phi|$ is small so that the binomial approximation to the distribution of the number of axis crossings is adequate, it follows directly that

$$\sqrt{n}(\hat{\phi}-\phi) \xrightarrow{L} N(0,\pi^2\lambda_1(1-\lambda_1)), \quad n \to \infty.$$

Estimation of the mean frequency. Let $Z(t)$ $-\infty < t < \infty$, be a zero mean stationary Gaussian process with correlation function $\rho(t)$. Assume that the sample functions are continuous with probability one and that in a sufficiently small time interval, say Δ, the probability that $Z(t)$ has more than one 0 is negligible. Consider the interval $[0,t]$ and partition it into $(n-1)$ subintervals of size Δ. Then $(n-1)\Delta = T$. We hold T fixed as $\Delta \to 0$ and $n \to \infty$ simultaneously. Let $X_{i,n}$ take the value 1 when $Z((i-1)\Delta) \geq 0$, and 0 otherwise, $i = 1,\ldots,n$, and let $D(n)$ be the number of symbol changes in the $X_{i,n}$ series. If D is the true number of axis crossings in $[0,T)$ then $D(n) \to D$, $n \to \infty (\Delta \to 0)$ a.s. As a <u>first</u> approximation to the distribution of $D(n)$ we take $D(n) \sim b(n-1,\ 1-\lambda_{1,n})$. Thus, by l'Hospital's rule

$$E(D) = \lim_{n\to\infty} E(D(n)) = \lim_{\Delta\to 0} \frac{T}{\Delta}\left[\frac{1}{2} - \frac{1}{\pi}\,\mathrm{si}^{-1}(\rho(\Delta))\right] = \frac{T}{\pi}\,\gamma,$$

$$\gamma = [-\rho''(0)]^{1/2}$$

provided the derivative exists. γ is called the mean frequency. A reasonable estimate for γ is then [4]

$$\hat{\gamma} = \frac{\pi D(n)}{T},$$

whose approximate distribution is easily obtained from that of $D(n)$.

References

[1] Kedem, B. (1976,a), Sufficient statistics associated with a two-state second-order Markov chain. <u>Biometrika</u>, 63, 127-132.

[2] Kedem, B. (1976,b), Exact maximum likelihood estimation of th parameter in the AR(1) process after hard limiting. <u>IEEE. TRANS. Info. Theory</u>, IT-22, 491-493.

[3] Leadbetter, M.R. (1972). Point processes generated by level crossings. <u>Stocahstic Point Processes</u>, A.W. Lewis, Editor, 436-467, Wiley, New York.

[4] Lindgren, G. (1974), Spectral moment estimation by means of level crossings. <u>Biometrika</u>, 61, 401-418.

This research was supported by the U.S. Air Force Office of Scientific Research Grant AFOSR 77-3332.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER (18) AFOSR-TR- 77-1259 (19)	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) SUFFICIENCY AND THE NUMBER OF LEVEL CROSSINGS BY A STATIONARY PROCESS.		5. TYPE OF REPORT & PERIOD COVERED (9) Interim rept.
		6. PERFORMING ORG. REPORT NUMBER (14) TR-77-41
7. AUTHOR(s) (10) Benjamin Kedem		8. CONTRACT OR GRANT NUMBER(s) (15) AFOSR-77-3332
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of Maryland Department of Mathematics College Park, MD 20742		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (16) (17) 61102F 2304 A5
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM Bolling AFB, DC 20332		12. REPORT DATE (11) August 1977
		13. NUMBER OF PAGES 8
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) (12) 10p.		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

binary, Markov chain, high levels, Poisson

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

It is shown how to derive the exact distribution of the number of axis crossings by a stationary process when the binary process obtained by clipping the original process is a pth-order Markov chain. The same method is used in deriving the asymptotic distribution of the number of upcrossings of a high level by a stationary process.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

400 953